DOCUMENT RESUME

ED 175 815 SP 014 197

AUTHOR Fox, Charles D.; Tubb, Gary W.
TITLE An Item Analysis of Teacher Evaluation Forms as a Mandate of Legislative Accountability.
PUB DATE Feb 79
NOTE 30p.; Paper presented at the Annual Conference of the Association of Teacher Educators (Orlando, Florida, February 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Accountability; Efficiency; *Evaluation Criteria; *Evaluation Methods; Faculty Evaluation; Feedback; *Item Analysis; Measurement Techniques; Performance Criteria; Reliability; *Teacher Evaluation; Validity

ABSTRACT

Before teacher evaluation forms can have any significance for educational research their validity and reliability must be established. This study concerns itself with defining the suitability of the criteria-establishing items that may be contained on teacher evaluations. The areas of inquiry are: the rater (evaluator) difficulty level of each item; the extent to which each rater discriminates between good and poor items; and the establishment of internal and external consistency in rating. A sampling of five administrators from a public school system in Louisiana brought variable results regarding the consistency levels in these areas. The results are included in statistical form in the appendixes to the report. (LH)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED175815

TITLE: An Item Analysis of Teacher Evaluation Forms as a Mandate of Legislative Accountability

PRESENTERS: Charles D. Fox, Winnfield Senior High School, Winnfield, LA.
Gary W. Tubb, Northwestern State University

ABSTRACT

In many states, accountability has been implemented by the state legislatures. Subsequently, various instruments have been developed for the purpose of measuring the accountability of teachers. Teacher evaluation forms have primarily been used as the sole standard for this measurement. Non-educators (i.e., legislators) have mandated criteria for measuring accountability through teacher evaluation forms. The purpose of this paper is to present a scientifically developed procedure for analyzing teacher evaluation forms. Item Analysis will be utilized in developing this procedure.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Gary W. Tubb
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at the annual National conference of the Association of Teacher Educators in Orlando, FL, February, 1979

SP 014 197

# ITEM ANALYSIS OF TEACHER EVALUATION FORMS

In its 1976 session, the Louisiana Legislature enacted legislation with far-reaching implications for all of Louisiana's public schools. This legislation, Act 709, provided for and described an accountability program which was to be practiced in the public schools.

During the 1976-77 school year, twenty-six of the sixty-six school districts in Louisiana initiated accountability programs in their systems. In March, 1977, the State Department of Education issued a guide for the implementation of Act 709, thus allowing the other districts to begin planning their accountability programs.

Reading and mathematics proficiency tests were administered in grades three, seven, and ten during the 1976-77 school year. The information derived from these tests were used in the development of criterion-referenced tests that were administered in all fourth, eighth, and eleventh grades during the 1977-78 school year.

The accountability act also provided for the evaluation of the performance of certified school personnel. In accordance with Act 621, enacted by the 1977 session of the Louisiana Legislature, evaluation forms used to assess teacher performance were developed by each local school district. (See Appendix A) (Louisiana State Department of Education, 1977:2-3)

Locally prepared teacher evaluation forms have been used as evaluative instruments by untrained evaluators or raters for many years. When utilizing these forms, some method must be used to determine the reliability and validity of these instruments. Familiarity with the basic concepts and techniques of item analysis can help the researcher

in his evaluation of these instruments. With item analysis, items on a teacher evaluation form can be analyzed for rater difficulty and inter-rater reliability.

In the past there was a tendency that any teacher evaluation form could be abused, especially when used by untrained raters. Various items on a form may be too difficult to rate or there may be little reliability between raters. High reliability and validity are very important for a teacher evaluation form to be of worth not only to the rater but also to meet the purpose set by administrators. Through item analysis, high reliability and validity can be built into a teacher evaluation form. These forms, therefore, can be improved by the selection, substitution, or revision of items. (Anastasi, 1976:198)

## PROBLEM

Each local public school system was required to develop programs for the evaluation of teachers and principals within its system as mandated by Act 9, passed by the 1977 session of the Louisiana Legislature. Once these programs received approval from the State Department of Education, the local school systems were to implement the evaluation programs beginning with the 1978-79 school year. However, training was not provided in the Act for those persons that would be evaluating teachers and principals. (See Appendix B) (Louisiana State Department of Education, 1978:1)

Before teacher evaluation forms have any meaning for use by administrators, the reliability and validity of the instrument should be established. One of the basic concerns of the present investigation was to determine the reliability and validity of teacher evaluation

forms intended for use by untrained raters. Louisiana's accountability law prescribed that teacher evaluation forms be employed to measure teacher effectiveness. However, without training of raters, the reliability of the instrument is questionable. Specifically, this study was concerned with three main questions:

1. What is the rater difficulty level of each item?
2. To what extent does each rater discriminate between good and poor items?
3. Can internal and external consistency be obtained in rating?

## METHOD

### Subjects

Five administrators from Winn Parish were randomly selected for the sample. Only those administrators who evaluate teachers were chosen. Personal and occupational characteristics of these administrators include the following: four male and one female; two principals, two assistant principals, and one supervisor of elementary education.

### Procedure

After reviewing twenty-seven of Louisiana's sixty-six teacher evaluation forms, it was found that each form contained the item: "teacher's knowledge in his/her subject area." Regardless of the particular rating system, this item was found on each form.

In order to determine rater difficulty in rating this item, the researcher chose the behavioral evidences of Standard Two (Demonstrates Wide and Growing Knowledge in Subject Area). (See Appendix C) Standard Two is found in Professional Competencies and Qualities of A Manual

for Developing Reasonable, Objective, Nondiscriminatory Standards for Evaluating Teacher Performance, published by the South Carolina Department of Education. (1974:11-12) These six behavioral evidences were given to each rater. The raters were then asked to think of the ideal teacher and rate each item from one to six. The behavioral evidence most reflective of a teacher's knowledge in his/her subject area would be given a rating of one. A rating of six would indicate that the item was least reflective of a teacher's knowledge in his/her subject area.

A Coefficient of Concordance was then applied in order to determine whether any agreement among the raters existed. Consequently, the more agreement among the raters on an item indicated a lesser difficulty level for the raters in rating the item.

To determine the discriminating power of the item found on teacher evaluation forms, the same behavioral evidences were used. This time, the behavioral evidences were paired so that each behavioral evidence was paired with the other five. Thus, fifteen pairs resulted. (See Appendix D) Again, with the ideal teacher in mind, each rater was asked to select one of the items in each pair which was most reflective of a teacher's knowledge in his/her subject area. A Coefficient of Consistence was then applied in order to determine if the task required a refinement of discrimination which was beyond the capacity of the raters.

## RESULTS

### Statistical Analysis of Item Difficulty

In order to analyze rater difficulty in rating an item, a Coefficient of Concordance was applied. (See Appendix E) According

to the findings of this statistical technique, W = .433 which was significant at the .057 level. This data shows that the raters were in agreement. Rater A, however, disagreed with the others on items a and d.

Statistical Analysis of Item Discrimination

A Coefficient of Consistence was applied to analyze item discrimination. Rater A had K = .375 and d = 5. In other words, Rater A was inconsistent and had a total of five inconsistent triads. Raters B, C, D, and E had K = 1 and d = 0. Essentially, the remaining four raters were consistent and had no inconsistent triads. (See Appendix F) Although inconsistency can be a problem in the rating of teachers, four of the selected raters were quite consistent.

Statistical Analysis of Internal and External Consistency

In order to determine internal and external consistency the following table is presented:

Table 1

Internal and External Consistency of Raters

| Rater | Type of Rating | Rating |
|---|---|---|
| A | Triad | a > b > c > d > a > e > f > b |
| | Concordance | d > b > a > c > f > e |
| B | Triad | a > f > b > c > e > d |
| | Concordance | a > f > b > c > e > d |
| C | Triad | a > c > b > e > f > d |
| | Concordance | a > c > b > e > f > d |
| D | Triad | a > d > f > b > c > e |
| | Concordance | a > f > d > b > e > c |
| E | Triad | a > c > f > b > c > d |
| | Concordance | a > c > f > b > e > d |

The symbol > means that the rater chose the first behavioral evidence listed over the second behavioral evidence. Triad rows resulted from the Coefficient of Consistence and the concordance rows resulted from the Coefficient of Concordance.

Results from Table 1 show that raters B, C, D, and E had internal and external consistency in rating teachers. In other words, these raters agreed with themselves and with others. Rater A did not agree with himself or with others.

## DISCUSSION

The item analysis of the item, "teacher's knowledge in his/her subject area," shows that this item has little rater difficulty, but high rater discrimination. As far as rater difficulty was concerned, this item was not very difficult to rate. Also, this item had a relatively high discriminating power.

This method of item analysis also provides an employer with a means by which to employ persons to rate teachers. Thus, the employer can establish that the rater has internal and external consistency in rating, with the result being a more consistent teacher evaluation inspite of untrained raters.

BIBLIOGRAPHY

Anastasi, Anne. Psychological Testing. New York: Macmillian Publishing Co., Inc., 1976.

Louisiana State Department of Education. "Accountability: A Guide for Implementation," 1977, 2-3.

Louisiana State Department of Education. "Accountability: Principal-Teacher Evaluation," 1978, 1.

South Carolina Department of Education. "A Manual for Developing Reasonable, Objective, Nondiscriminatory Standards for Evaluating Teacher Performance," 1974, 11-12.

APPENDIX A

School Personnel Assessment and Evaluation. The SDE, in cooperation with the Task Force on Personnel Assessment and Evaluation, shall develop a mutually acceptable set of guidelines for assessment and evaluation of the performance of certified school personnel. Initial compliance will yield various evaluation procedures and criteria. Subsequent refinement by local and state committees will yield minimal statewide evaluation guidelines for adoption by each local board.

Requirements for Guidelines as specified by Act 621:

1. Each LEA shall develop a system based on established criteria for screening new applicants for positions requiring certified school personnel.
2. All evaluation systems shall provide for observations, conferences, written recommendations for improvement, and a copy of the evaluation to the employee within 15 working days after said evaluation.
3. Persons evaluated shall have the right to initiate a written reaction or response to the evaluation. Such response and evaluation shall become a permanent attachment to the single official personnel file for the employee.
4. Evaluations shall be made on a continuing basis at least once each school year for probationary personnel and at least once every third year for personnel with permanent status. When the performance of permanent status personnel has been evaluated as unsatisfactory, subsequent evaluation will be conducted on a yearly basis until identified deficiencies have been corrected.
5. All evaluation procedures shall include the use of an evaluation form.

6. All evaluation forms shall be countersigned by the person evaluated under a disclaimer clause.
7. Any employee who has been evaluated and requests a meeting with the appropriate official concerning said evaluation shall be granted said meeting in order that the employee may respond to the evaluation.
8. Evaluation forms and related material shall be kept on file within the LEA, attached to the single official personnel file for the employee and are to be confidential, except as prescribed by law.
9. All evaluation procedures shall include an in-service component which shall have professional growth as one of its primary objectives.
10. All actions resulting from the evaluations shall take into account the principles of due process.

In meeting the above requirements, LEAs are mandated by Act 621 to be guided by the following provisions:

1. The SDE, in conjunction with advisory committees, shall develop, prior to April 15, 1978, guidelines for LEA certified personnel evaluation procedures, such procedures to be adopted by the Superintendent of Education prior to June 1, 1978.
2. By August 15, 1978, LEAs shall submit to the SDE a locally approved plan which shall reflect the above guidelines.
3. LEAs shall implement local evaluation procedures for certified school personnel by September 1978.

4. LEAs shall complete criteria for expected performance for all certified school personnel by July 1979. Job descriptions will be an outgrowth of this process.
5. During the 1979-80 school year, the SDE shall conduct in-service programs to assist the LEAs with the implementation of state adopted personnel evaluation guidelines.
6. Full implementation of an approved evaluation plan for all certified school personnel must be in effect by August 1980.

A report of the evaluation process shall be submitted by the LEAs to the SDE by June 15 of each year. The SDE shall submit a summary report to the Joint Legislative Committee on Education by September 1 of the same year.

## APPENDIX B

1. To provide for programs for evaluation of teachers and principals in the parish and city school systems of the state.

2. To require the local school boards to submit plans of evaluation programs to the SDE in consonance with SDE guidelines for approval.

3. To provide for the establishment of guidelines and procedural rules by the SDE. The resulting guidelines will be in conjunction with the efforts of the SDE and the 66 member State Task Force on Personnel Evaluation.

4. To require technical assistance to the local boards by the SDE in developing evaluation programs.

5. To require that teachers and principals found in need of professional improvement participate in special training to upgrade professional competencies.

## APPENDIX C

____Demonstrates wide knowledge in subject or grade level and related fields;
____Relates subject to current issues and problems;
____Encourages critical thinking and examination of all sides of all questions;
____Freely admits he does not know all answers and demonstrates that he is a learning individual also;
____Explains popular misconceptions;
____Relates understandings to assigned areas and objectives.

APPENDIX ?

For each pair of questions select the one most reflective of a teacher's knowledge in his/her subject area.

1. a___Demonstrates wide knowledge in subject or grade level and related fields.
   b___Relates subject to current issues and problems.

2. a___Encourages critical thinking and examination of all sides of all questions.
   b___Freely admits he does not know all answers and demonstrates that he is a learning individual also.

3. a___Explains popular misconceptions.
   b___Relates understandings to assigned areas and objectives.

4. a___Demonstrates wide knowledge in subject or grade level and related fields.
   b___Freely admits he does not know all answers and demonstrates that he is a learning individual also.

5. a___Relates subject to current issues and problems.
   b___Explains popular misconceptions.

6. a___Encourages critical thinking and examination of all sides of all questions.
   b___Relates understandings to assigned areas and objectives.

7. a___Demonstrates wide knowledge in subject or grade level and related fields.
   b___Explains popular misconceptions.

8. a___Relates subject to current issues and problems.
   b___Relates understandings to assigned areas and objectives.

9. a___Demonstrates wide knowledge in subject or grade level and related fields.
   b___Encourages critical thinking and examination of all sides of all questions.

10. a___Relates subject to current issues and problems.
    b___Freely admits he does not know all answers and demonstrates that he is a learning individual also.

11. a___Encourages critical thinking and examination of all sides of all questions.
    b___Explains popular misconceptions.

12. a___Freely admits he does not know all answers and demonstrates that he is a learning individual also.
    b___Relates understandings to assigned areas and objectives.

13. a___Relates subject to current issues and problems.
    b___Encourages critical thinking and examination of all sides of all questions.

14. a___Demonstrates wide knowledge in subject or grade level and related fields.
    b___Relates understandings to assigned areas and objectives.

15. a___Freely admits he does not know all answers and demonstrates that he is a learning individual also.
    b___Explains popular misconceptions.

APPENDIX E

Coefficient of Condorance

| Administrator | Definition | | | | | |
|---|---|---|---|---|---|---|
| | a | b | c | d | e | f |
| A | 3 | 2 | 4 | 1 | 6 | 5 |
| B | 1 | 3 | 4 | 6 | 5 | 2 |
| C | 1 | 3 | 2 | 6 | 4 | 5 |
| D | 1 | 4 | 6 | 3 | 5 | 2 |
| E | 1 | 4 | 2 | 6 | 5 | 3 |
| $R_j$ | 7 | 16 | 18 | 22 | 25 | 17 |

W = .433

APPENDIX F

Coefficient of Consistence - Rater A

| | a | b | c | d | e | f | R | $(R - \bar{R})^2$ |
|---|---|---|---|---|---|---|---|---|
| a | . | 1 | 0 | 0 | 1 | 1 | 3 | .25 |
| b | 0 | . | 1 | 1 | 1 | 0 | 3 | .25 |
| c | 1 | 0 | . | 1 | 1 | 1 | 4 | 2.25 |
| d | 1 | 0 | 0 | . | 1 | 1 | 3 | .25 |
| e | 0 | 0 | 0 | 0 | . | 1 | 1 | 2.25 |
| f | 0 | 1 | 0 | 0 | 0 | . | 1 | 2.25 |

K = .375 d = 5

Note: 1 means that the item in the row was chosen over the item in the column.
0 means that the item in the row was not chosen over the item in the column.

Coefficient of Consistence - Rater B

| | a | b | c | d | e | f | R | $(R - \bar{R})^2$ |
|---|---|---|---|---|---|---|---|---|
| a | . | 1 | 1 | 1 | 1 | 1 | 5 | 6.25 |
| b | 0 | . | 1 | 1 | 1 | 0 | 3 | .25 |
| c | 0 | 0 | . | 1 | 1 | 0 | 2 | .25 |
| d | 0 | 0 | 0 | . | 0 | 0 | 0 | 6.25 |
| e | 0 | 0 | 0 | 1 | . | 0 | 1 | 2.25 |
| f | 0 | 1 | 1 | 1 | 1 | . | 4 | 2.25 |

K = 1 d = 0

Coefficient of Consistence - Rater C

| | a | b | c | d | e | f | R | $(R - \bar{R})^2$ |
|---|---|---|---|---|---|---|---|---|
| a | . | 1 | 1 | 1 | 1 | 1 | 5 | 6.25 |
| b | 0 | . | 0 | 1 | 1 | 1 | 3 | .25 |
| c | 0 | 1 | . | 1 | 1 | 1 | 4 | 2.25 |
| d | 0 | 0 | 0 | . | 0 | 0 | 0 | 6.25 |
| e | 0 | 0 | 0 | 1 | . | 1 | 2 | .25 |
| f | 0 | 0 | 0 | 1 | 0 | . | 1 | 2.25 |

K = 1 d = 0

Coefficient of Consistence - Rater D

| | a | b | c | d | e | f | R | $(R - R)^2$ |
|---|---|---|---|---|---|---|---|---|
| a | . | 1 | 1 | 1 | 1 | 1 | 5 | 6.25 |
| b | 0 | . | 1 | 0 | 1 | 0 | 2 | .25 |
| c | 0 | 0 | . | 0 | 1 | 0 | 1 | 2.25 |
| d | 0 | 1 | 1 | . | 1 | 1 | 4 | 2.25 |
| e | 0 | 0 | 0 | 0 | . | 0 | 0 | 6.25 |
| f | 0 | 1 | 1 | 0 | 1 | . | 3 | .25 |

K = 1 d = 0

Coefficient of Consistence - Rater E

| | a | b | c | d | e | f | R | $(R - R)^2$ |
|---|---|---|---|---|---|---|---|---|
| a | . | 1 | 1 | 1 | 1 | 1 | 5 | 6.25 |
| b | 0 | . | 0 | 1 | 1 | 0 | 2 | .25 |
| c | 0 | 1 | . | 1 | 1 | 1 | 4 | 2.25 |
| d | 0 | 0 | 0 | . | 0 | 0 | 0 | 6.25 |
| e | 0 | 0 | 0 | 1 | . | 0 | 1 | 2.25 |
| f | 0 | 1 | 0 | 1 | 1 | . | 3 | .25 |

K = 1 d = 0